smartKYC

Overcoming KYC Challenges

implementing a continuous or high frequency KYC monitoring programme

March 2021

Contents

Introduction 1
1. Decide what you want to monitor 1
2. The inconvenience of languages 2
3. Managing content costs 2
4. Déjà vu and avoiding repetition 3
5. Reducing external search footprint 4
6. False positive minimisation 4
7. Source authority 5
8. Accentuate the positive 5
9. Joined up thinking 6
10. Consume media responsibly 6

Introduction

In 2004, a colleague and I had the idea of building a search string of negative words in several languages and applying that to a corpus of news content to support our banking clients with their KYC efforts. Thus was born the world's first adverse media search tool, Nexis Diligence, and quite some time before the FSA (now Financial Conduct Authority) recommended the use of such tools for KYC.

Today, and now running a KYC technology business, I think it is inevitable that best practice guidance will soon enough be looking to add high frequency monitoring to the two existing KYC pillars of activity, onboarding due diligence and periodic refresh.

But many are daunted by the prospect of implementing a high frequency KYC monitoring programme – the prohibitive capital costs, the huge increase in operating expenditure to manage the increased workload, the nightmare scenario of creating another transaction monitoring-type SARs headache: it's hard to know where to start.

Despite the challenges, we sense an increased appetite by banks to stay one step ahead of the regulator by implementing such a programme. So, with the benefit of the last 15 years' experience in the field, let's make a start on the challenges and how to overcome them.

1. Decide what you want to monitor

While it's perfectly possible and perhaps desirable to look for any change to the intelligence gathered at the onboarding stage, is it really necessary to watch all clients so proactively and across multiple information parameters? Some, like country or sector exposure, environmental, social and governance (ESG) breaches or even a source of wealth development might not require daily alerting and can be picked up at the next periodic refresh cycle. Banks should develop a framework for what KYC looks like for each of these three activities.

So while onboarding KYC could include screening for all risks plus consideration of contextual intelligence (background narrative, risk by association, journey to wealth and corroboration, etc.), high frequency monitoring might focus on breaking, adverse legal reporting only.

But perhaps go one step further and define adverse materiality in order to avoid overload. The system you implement should be both highly granular in its classification of risk and nuanced enough to distinguish between type of legal matter (e.g. is a parking offence of sufficient gravity to merit an alert?) It should also be able to understand the context of awkward words like 'accusation'. There is a world of difference between being accused by the International Criminal Court and being accused of infidelity. An automated semantic search and content analysis platform should make these distinctions for you.

2. The inconvenience of languages

It goes without saying that if there is an international element to a bank's client base, whether that's birth, domicile or footprint, monitoring English language content alone will obviously expose the bank to the risk of missing important intelligence. It stands to reason that extracting intelligence with precision from foreign language sources is a must. However, some approaches machine-translate non-English material into English first and then derive results based on the translation. With such an approach, 'lost in translation' doesn't just mean loss of nuance but potentially and critically, a loss of original meaning.

1. مصالح الدرك تحقق مع اسعد ربراب

   A Google translation of this Arabic sentence is: "The interests of the gendarmerie were achieved with Asaad Rabrab."
   What it really means is: "The gendarmerie investigates Issad Rebrab." The key term 'investigate' is lost and a potential risk is not flagged up.

2. **黎明从未成年起就抢劫伤人，二进宫以后仍然不悔改**

   Similarly, this Simplified Chinese excerpt is machine translated as: "Dawn has never robbed and wounded people since he was an adult."
   Whereas in English it means: "Li Ming has started robbing and hurting people since he was a juvenile, and never repented after he was re-incarcerated ."

3. It's not just those non-Latin languages where these mistakes can be found. Problems can arise closer to home. **"Attaqué en diffamation par Cédric Herrou, Eric Ciotti relaxé."**
   Machine translates as: "Defamed by Cédric Herrou, Eric Ciotti released."
   But the actual meaning is the opposite: "Accused of defamation by Cédric Herrou, Eric Ciotti is released."

And isn't it ironic that in our KYC / AML world, one of the most familiar words, 'sanction', makes the case for a semantic approach as it can mean opposing things, depending on whether it is a noun or a verb?

These apparently minor mistakes could have major consequences. Stripping away the syntactical, grammatical and lexical rules of the original language of the source material for the sake of the 'convenience' of processing everything in English will prove an expensive shortcut.

3. Managing content costs

A common fear is that the content cost of searching many thousands of client names via premium media sources will prove prohibitive. But this is not a look-back exercise and therefore access to a complete news archive of premium licensed materials is wholly unnecessary. Content providers have solutions better suited to monitoring the last 24-hour news cycle only, which are both cost-effective (i.e. predominantly unlicensed and an unmetered price tag) and of excellent quality.

There is also an emerging alternative to the above inside-out approach (i.e. these are my client names, search them every day / hour) and that is to do it in reverse. The outside-in approach processes the last 24 hours news cycle for adverse references, extracts the actors from the articles and serves this as a feed to the bank. Matching to the bank's client names can either be a manual process or by using sophisticated name matching tools.

4. Déjà vu and avoiding repetition

In a previous life as a director of the world's largest media monitoring group, my primary audience was corporate communications professionals. The parallel with continuous KYC monitoring is obvious in that both are interested in what is being reported in the media but there is one essential difference. PRs love the repetition of a story as it constitutes success. For KYC, repetition is the enemy as it grinds the gears of the KYC operation. And in my 30 years of leadership roles in representing all sides of news media – a national newspaper, an aggregated news archive, a news distributor, a media monitoring company – one thing is for sure, news repetition is more prevalent than ever. This is due partly to the economics of newspapers (content syndication, the selling of republishing rights, is an increasingly important revenue line for newspapers) and also due to the proliferation of online news aggregators which means content is regurgitated like never before.

Minimising duplicate articles is technically pretty simple – essentially it is basic plagiarism detection, which, in essence, is searching for the same 10 or so words in a row. Far more challenging is to identify what we call 'informationally similar' content i.e., this is the same story but from a different source using different wording, maybe in an altogether different language. Through the automatic semantic extraction of common facts from seemingly disparate documents, an intelligence baseline is established such that the bank can be alerted only to genuinely new 'facts' in future rather than sit in a metaphorical echo chamber, reading the same stuff over and over.

Consider the following three separate media snippets when screening the company ZTE:

- "Регулятор объяснил свое решение тем, что ZTE, которая была оштрафована властями США на $1,2 млрд за нарушение санкционного режима в отношении Ирана и Северной Кореи, нарушила условия соглашения, заключенного для урегулирования этого дела."

- "Há três anos, a ZTE se declarou culpada por violar sanções comerciais contra o Irã e a Coreia do Norte."

- "In March 2017, ZTE agreed to pay a record fine of $1.2 billion to the United States for violating sanctions on Iran and North Korea."

These three snippets are written completely differently, in different publications, languages and, in one instance, a different alphabet. They all however pertain to the same fact that this company was violating sanctions.

An analyst shouldn't have to read all three but without the technological means to aggregate these and the many other media references reporting this same fact, they may have no option but to wade through them all.

5. Reducing external search footprint

For varying security reasons, some institutions are sensitive to the volume of searches leaving their environment. Potentially this issue would be exacerbated with the implementation of a high-frequency monitoring programme, but minimisation of external search traffic can be easily achieved by:

a) Hosting the base media content on-premise with updates received daily, so that the search doesn't leave the bank's environment (content providers are amenable to this deployment model)
b) Using specialist search anonymisation techniques
c) Considering the outside-in approach, in point 3 above

6. False positive minimisation

In point 2 we saw how a sophisticated approach to languages and delivering genuine news deltas through fact extraction (point 4) help minimise false hits and repetitious noise. But arguably the main contributor to work inefficiency and general frustration is false positives due to mistaken identity – going down KYC research rabbit holes only to establish that what you are reading is not really about your search target.

One misapprehension is that adverse media searching is solely about looking for a name in proximity to an adverse term and sifting through the results. But news items will often yield richer information from which technology can 'disambiguate' identity. News articles will often contain the name, age, nationality, area of residence, and sometimes occupation, if not employer, of the subject of the article. Semantic search technology can understand these various textual elements for what they are and map them to the identifying attributes that were captured at onboarding. That way a confidence score can be ascribed to a hit based on its likelihood to be about the bank's client. But rest assured that for security reasons these identifying attributes do not need to leave the building as all the processing should be done on premise.

And while on the subject of names, do ensure that where appropriate, the name of your KYC subject is transliterated (i.e. searched in its original form or alternate form), ideally automatically to avoid human error. A search for a well-known, alleged Russian criminal yields ten times the results using the original Cyrillic version of the name compared to the Latin equivalent (the fact that you are presented with less than 0.02% of this material by a well-known search engine is quite another matter). Not that you will need to read all of them anyway, your search tool should do all that heavy-lifting for you. The point is that as with human source intelligence gathering, a local language perspective significantly reduces the risk of missing a vital piece of information.

7. Source authority

This is a particularly vexing issue. Banks, quite understandably, value story veracity so that they don't act on unreliable or, worse, fake news. But that often leads to the inclination to rely only on a small pool of 'trusted' media sources, a national media outlet for example. But, a balance needs to be struck here:

a) There are now many credible, trustworthy online news sites and blogs whose output is only available directly, not via a third party. Many of these have sophisticated news gathering operations and will be particularly tuned into their local market. In fact, they are often used by other bigger international media outlets as a primary news source. It would be foolhardy to ignore them.
b) Waiting for a 'reputable' source to report on something has two risks: the first is latency. Stories break somewhere and can take a long time to percolate up, especially if it happens in a foreign jurisdiction. The second is audience-relativity. Unless the subject is high profile, most stories won't make it past the copy taster's desk of that national news outlet (the copy taster being the person at a news outlet whose job it is to decide what might be newsworthy). This is especially so if the story relates to a foreign national given the domestic agenda of most news outlets.
c) It is also worth rethinking your definition of 'intelligence' and its place in KYC monitoring. We can point to an example where the first suggestion of political corruption in a very high profile case was actually made in the comments section of a celebrity gossip site. Such comments are accessible only via search engines but should they be ignored as they lack authority? Should the site be dismissed altogether?

It is perhaps worth pointing out that by subscribing to a paid-for media archive you are not only accessing historic content that mightn't be available elsewhere (due to publication date, paywall or licensing restrictions). These providers can rightfully argue that given they curate and pay for aggregating news sources, there is, de facto, source authority by design.

8. Accentuate the positive

We use the phrase 'relationship intelligence' because this needn't be just about risk. KYC shouldn't be the exclusive preserve of the compliance function. I started my career in sales and it was drummed into me that your 'edge' was gained by knowing as much about your target as possible; background, hobbies, the nature and direction of a business, etc.

Quite rightly compliance will say that extracting the positives isn't our responsibility and besides we already have enough on our plate. But the business might need to think about a more holistic approach as looking for positive events and indicators is just the flip side of the same KYC coin. The business, not necessarily compliance, could be watching for opportunity as well as risk using the same technology, tailored to the needs of the business line – a liquidity event, an asset disposal, a personal event, a specific corporate action…
Such things are understandably far from the minds of financial crime professionals but positioned as part of a broader client lifecycle management play, the monitoring investment argument may be given additional heft and not seen as just another grudge compliance purchase.

9. Joined up thinking

A potential pitfall is to think that while these three KYC activities (onboarding due diligence, periodic refresh and high frequency monitoring) are distinct, there is no need to treat them as intrinsically linked. It makes no sense to establish an intelligence baseline at onboarding with one system but monitor that client relationship using another – and this is not just because there will be two discrete systems to learn and support. Not only will there be different rubrics to how the systems approach their separate tasks but also, that factual baseline established at onboarding cannot be used to ensure monitoring noise is kept to a minimum (see Point 4 of this article, Déjà vu and avoiding repetition).

A key factor for any KYC implementation is auditability and one of the main criteria in the investment decision of the banks with whom we work. Any system must record not only what was searched and when but also what actions were taken, who took those actions, what material was regarded as relevant and what was discarded. It's perhaps not the most exciting aspect of what we do but imagine the alternative - manual logging by humans, errors, a mess of unstructured supporting documentation, all against a backdrop of human inconsistency.

Another consideration is how your KYC monitoring system 'talks' to your CLM (client lifecycle management) system because that solution is the one entrusted with the broader orchestration of the various aspects of the client relationship. Potential hits need to be routed correctly, triaged for further investigation, managed according to necessary roles and permissions, outcomes appropriately recorded and synchronicity maintained between monitoring system and CLM. Your provider needs to have the requisite technology and professional skills to understand these requirements in tedious detail and implement them.

10. Consume media responsibly

At the time of writing I am unaware of a licence that allows an organisation carte blanche to consume any media content in whichever way it chooses. Even if such existed, I'm not sure it would explain some common practices, most likely 'belt and braces' measures such as downloading page after page of materials, taking endless screengrabs of results.

Notwithstanding the fact that media providers have historically been known to collectively bear their teeth re alleged media monitoring malpractice (matters have reached the UK Supreme Court, notably the Newspaper Licensing Association v Meltwater case), from a practical perspective surely it doesn't make sense to hoover up disparate and often extraneous content in varying formats in the interest of due diligence. Far better to search with a higher degree of precision and generate a consolidated, economical and readily portable dossier that is also consistent, review to review and from onboarding, to refresh, to monitoring.

Meeting the ten challenges of implementing an effective high frequency KYC monitoring programme may be a tall order but it promises the ultimate compliance defence – eyes always open. Success requires a combination of precision technology, deep KYC domain knowledge and an ability to navigate the increasingly complex world of mainstream and social media.